# MSDS 413, Summer 2019, Assignment 1
# Forecasting - Fundamental Concepts (TS1)

## Introduction

For this assignment, you will use the dataset m-ge3dx8113.txt, which was included in the Session 1 zip file (TS1.zip) and, along with a R script (TS1.R), is included on Canvas. You will use the Australian wine sales data is in AustralianWine.csv, bit coin price data in bitcoin_cash_price.csv, and the visitors numbers (in expsmooth via fpp packakge). You will read the files into R or other statistics package and conduct the requested analyses.

The following list defines the variables:

- AustralianWine.csv (volumes are 1000× Kiloliters
  - Month
  - Fortified
  - Red
  - Rose
  - sparkling
  - Sweet.white
  - Dry.white
- bitcoin_cash_price.csv
  - Date
  - Open
  - High
  - Low
  - Close
  - Volume
  - Market.Cap
- visitors numbers (in expsmooth via fpp packakge) as an array of
  - year (beginning May 1985 - April 2005)
  - month (derive from case sequence as per year)
  - visitor quantity

Your objective is to explore the relationship among the variables, identify what types of data each variable is, and respond to each part in the Procedure Section that follows.

## Procedure

The following steps are necessary to complete this assignment. Address each and every part and ensure that you cover all the details specified in the questions.

1. **EDA** (1 point) Consider the daily prices of bitcoin currency from August 2, 2017, to February 20, 2018. The data file is bitcoin_cash_price.csv with column names Date, Open, High, Low, Close, Volume and Market.Cap, respectively.

    1.1. For the bitcoin daily prices (High, Low, Close) and daily volume (Volume), compute the sample mean, standard deviation, skewness, excess kurtosis, minimum, and maximum of each series. What do these statistics tell you?

    1.2. For the bitcoin daily prices (High, Low, Close) and daily volume (Volume), use the log transform of the prices and volume. Compute the sample mean, standard deviation, skewness, excess kurtosis, minimum, and maximum of each log series. Interpret and compare to the simple price and volume statistics in part 1.1..

    1.3. Test the null hypothesis that the mean of the log closing price is zero. Interpret.

    1.4. What is the pairwise association between the log closing price and the log volume? Interpret.

    1.5. Obtain the empirical density plot and normal QQ plot of the daily log closing price and the log volume. Interpret these plots.

2. **Hypothesis tests** (1 point) Consider again the daily prices of bitcoin currency from August 2, 2017, to February 20, 2018. Perform the following and draw conclusions using the 5% significance level.

    2.1. Construct a 95% confidence interval for the daily log closing price of the bitcoin data.

    2.2. Test $H_0 : m_3 = 0$ versus $H_a : m_3 \neq 0$, where $m_3$ denotes the skewness of the price.

    2.3. Test $H_0 : K = 3$ versus $H_a : K \neq 3$, where $K$ denotes kurtosis.

3. **Time series models 1** (2 points) Use the monthly Australian short-term overseas visitors data, May 1985 - April 2005. (Data set: visitors) in the Hyndeman text.

    3.1. Make a time plot of the data and describe the main features of the series.

    3.2. Forecast the next two years using Holt-Winters' multiplicative method. Assess the forecast viablility.

    3.3. Why is multiplicative seasonality necessary here?

    3.4. Experiment with making the trend exponential and/or damped. Interpret and compare.

    3.5. Compare the RMSE of the one-step forecasts from the various methods. Which do you prefer and why?

4. **Time series models 2** (2 points). For this exercise, use the monthly Australian wine sales *Fortified* data. (Data set: AustralianWine.csv.)

4.1. Make a time plot of your data and describe the main features of the series

4.2. Forecast the next two years using Holt-Winters' multiplicative method.

4.3. Why is multiplicative seasonality necessary here?

4.4. Experiment with making the trend exponential and/or damped.

4.5. Compare the RMSE of the one-step forecasts from the various methods. Which do you prefer?

4.6. Now fit each of the following models to the same data:

4.6.1. a multiplicative Holt-Winters' method;

4.6.2. an ETS model;

4.6.3. an additive ETS model applied to a Box-Cox transformed series;

4.6.4. a seasonal naive method applied to the Box-Cox transformed series;

4.6.5. a STL decomposition applied to the Box-Cox transformed data followed by an ETS model applied to the seasonally adjusted (transformed) data.

4.7. For each model, look at the residual diagnostics and compare the forecasts for the next two years. Which do you prefer?

5. **Report** (1.5 point) For the Australian wine sales data set, provide an executive summary from which a CEO, e.g., can take action. Use the appropriate plots and statistics needed to support your recommendations. Keep it simple and to the point, no novels.

## Deliverables

See Section Submission Directions below. The assignment deliverables, each in pdf format, are as follows:

- *Only if requested by instructor*
  - The program or script
  - Logs
  - Outputs
- Mandatory
  Data analysis write-up: no programs, logs, or just code outputs.

The data analysis must follow and use the item numbering of each assignment, i.e., use the numbers, say, 1 - 5, with the sub-lettering if used. These deliverables are provided according to the instructions in the Submission Directions section below.

# Submission Directions

## Title Page

Include a title page with your name and the assignment designation. Leave room for instructor comments.

## File Names

The assignment write-up file shall be submitted to Canvas according to the schedule in the syllabus using the item (1) naming convention below. The naming convention is case sensitive. Use letters and numbers as given. **The file name parts have no spaces or other separator charancters**. TS1Lastname.pdf (submit via Canvas)

The parts are the assignment code, TS1; your lastname with only the first letter capitalized; a period, and lastly, the extension "pdf". Generically,

TS1Lastname.pdf

For example: Suppose your name is Student McStats. Your filename then is:

TS1Mcstats.pdf

The analysis write-up file must be submitted for grading. Each write-up requires a title page for instructor comments. The analysis may use either R or any other statistics package you wish, or if you use more than one package, you must use the germane tables, plots, etc., in a single report. If you use more than one package, differences and similarities should be indicated.

## email: jamie.riggs@northwestern.edu

Email *ONLY IF REQUESTED* the program (script), log and output as separate pdf files. The R log and output may be combined. The file names shall be as follows:

- The program or script file names
  - TS1LastnameRprog.pdf
- The log file names
  - TS1LastnameRlog.pdf